Total Variation and Signature-Based Regularizations on Coupled Nonnegative Matrix Factorization for Data Fusion

Yang, F.; Ma, F.; Ping, Z.; Xu, G.

Published in:
IEEE Access

DOI (link to publication from Publisher):
10.1109/ACCESS.2018.2879943

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 5, 2018, accepted October 7, 2018, date of publication November 9, 2018,
date of current version January 7, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2879943

Total Variation and Signature-Based Regularizations on Coupled Nonnegative Matrix Factorization for Data Fusion

FEIXIA YANG[1,2], FEI MA[3], ZILIANG PING[1], AND GUIXIAN XU [4]

[1]School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing 10076, China
[2]School of Electrical and Control Engineering, Liaoning Technical University, Huludao 125105,China
[3]School of Electronic and Information Engineering, Liaoning Technical University, Huludao 125105,China
[4]Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark

Corresponding author: Fei Ma (femircom@gmail.com)

This work was supported by the Basic Scientific Research Project of Colleges from the Liaoning Department of Education (P.R.C) under Grant LJ2017QL014.

ABSTRACT As an effective approach to gain the high-spatial-resolution hyperspectral images, data fusion is usually adopted to enhance the spatial resolution of hyperspectral images by the spatial information of multispectral images. In this paper, in order to remove the ill-posedness of well-known coupled non-negative matrix factorization, we formulate a well-posed fusion problem by incorporating total variation and signature-based regularizations for image smoothing and high-fidelity signature reconstruction. Then, the problem can be decoupled into two convex subproblems, which yield closed-form solutions separately by the alternating direction method of multipliers algorithms. Due to the large sizes of the problems, a few of constructed matrices and tensor operations are employed to simplify the expressions for reducing the computational complexities. Simulation and experimental results not only demonstrate that the performance of the proposed fusion algorithm is much better than that of state-of-the-art methods but also show that the total variation and signature-based regularizers are of paramount importance in yielding the high-spatial-resolution hyperspectral images.

INDEX TERMS Total variation, CNMF, data fusion, alternating direction method of multipliers.

I. INTRODUCTION

As we all know, several kinds of images in geoscience and remote sensing [1]–[4] are usually exploited, such as panchromatic (PAN), multispectral and hyperspectal images. Compared with multispectral images (MSIs), PAN images have higher spatial and lower spectral resolutions, while hyperspectal images (HSIs) usually have lower spatial and higher spectral resolutions. In practice, high-spatial-resolution HSIs are required for accurate identification and classification of the underlying materials (named as endmembers). Thus, how to obtain the high-spatial-resolution HSI data is an important and challenging topic in hyperspectral areas [5]. As an effective and promising approach, data fusion of low-spatial-resolution HSIs and high-spatial-resolution MSIs has been drawing considerable attentions to enhance the spatial resolution of HSIs [6]–[8]. There have been various methods based on different theories, so far, to achieve the goal of data fusion, which can be divided into three main categories including extended pan-sharpening methods, Bayesian-based methods and spectral unmixing-based methods.

Pan-sharpening, short for panchromatic sharpening [9]–[11], has been developed originally to fuse MSIs and PAN images for improving the spatial resolution of the MSIs. Corresponding to the feature relationships among HSIs, MSIs and PAN images, pan-sharpening can be thought of as the special case of the fusion of HSIs and MSIs [7]. In other words, pan-sharpening can be extended to the fusion of HSI and MSI data, whose representative methods mainly include component substitution [12]–[14] and multi-resolution analysis [15]–[17]. For instance, component-substitution-based pan-sharpening, originally substituting the component of MSIs by the corresponding part of PAN images, can be adapted to fuse the HSI and MSI data. The pansharpening methods based on multiresolution analysis [18] can be utilized by linearly combining multispectral band images to

synthesize a high-spatial-resolution image for each HSI band. However, pansharpening methods usually involve multiple spectral bands in data fusion, yielding significant spectral distortion and performance degradation.

Another popular approach is based on Bayesian probability. For example, a Bayesian method, using a maximum a posteriori (MAP) estimation and a stochastic mixing model [19], was proposed to enhance the spatial resolution of all hyperspectral band images in the principal component subspace. Wei *et al.* [20] proposed a sparse representation-based Bayesian method to address the ill-posed problem, by defining a proper posterior distribution based on a decomposition of the image in a set of dictionaries. A Sylvester-equation-based solution was integrated into the Bayesian fusion methodology, named as the fast fusion based on Sylvester equation [21], to drop the large computational complexity of Bayesian method [20]. As another representative, Simoes *et al.* [22] developed a method, termed as hyperspectral superresolution (HySure), by imposing the vector total variation regularizer on two convex data fitting terms for piecewise smoothing, which was solved via the split augmented Lagrangian shrinkage algorithm.

Another promising data fusion approach is known as coupled non-negative matrix factorization (CNMF) [23], which could not only yield the high-spatial-resolution HSIs, but also estimate spectral signatures and abundance distribution of the underlying materials in the meantime. Moreover, CNMF has a good property that only requires a larger model-order for the number of endmember instead of an accurate value as a prior knowledge [23]. However, CNMF criterion often amounts to an ill-posed inverse problem, such as when the number of MS bands is smaller than that of endmembers [7], it requires the regularization [24]. The most commonly used regularizers are to promote the sparsity of abundance matrix based on ℓ_0-norm or its convex surrogate ℓ_1-norm, and recently sparse regularization terms based on Frobenius norm have also been proposed [25]–[27]. However, the sparsity-promoting regularization itself may not be sufficient to yield high-quality fused data.

From the view of NMF, except for abundance regularizations, e.g. sparsity, the regularizations on spectral signature can be considered, which were derived from Craig's criterion on simplex volume minimization [28], [29]. To the best of our knowledge, the research work in this line was firstly proposed in [30], where the volume of signature vectors' simplex is employed to regularize the CNMF solution. However, the multiplicative update rules adopted in [30] can not yield an obvious enhancement of fusion performance. Recently, a state-of-the-art method, termed as convex optimization-based CNMF (CO-CNMF), was conducted by Lin *et al.* [31], where the signature regularization based on sum-of-squared distances between all the simplex vertices (i.e. the surrogate of the simplex volume), as well as sparsity-promoting term ℓ_1-norm, was employed to yield significant improvements of HSI-MSI fusion performance. However, further research suggests that the performance of CO-CNMF method would degrade rapidly with the increase of noise level, i.e., it is susceptible to high noise level. As a result, the image denoising or smoothing should be introduced into the fusion method. Total variation (TV) is a suitable term that has been extensively utilized for image removal and edge preserving, which has two forms including isotropic and anisotropic expressions [22].

In this article, to address the performance degradation in a high noise environment [31], we incorporate the anisotropic total variation and signature-based regularizations to reformulate the CNMF problem as a bi-convex problem. That is, it can be decoupled into two convex subproblems via alternating optimization, after converting the regularizations into the convex forms. Furthermore, we propose the carefully designed algorithms to yield closed-form solutions via ADMM, in which some equality constraints are employed to split the primal variables. However, these closed-form solutions have high computational complexity that are not suitable for fast iterative solvers. Therefore, some structured matrices and vector-matrix operators are exploited to greatly drop the heavy computational burden on the time-consuming ADMM iterations. In the experiments based on Wald's protocol [32], model-order selection is conducted to determinate the number of endmembers. The experimental results demonstrate that the proposed algorithm outperforms the baseline methods and enhance the quality of fused data, especially in low signal-to-noise ratio (SNR) environments.

This paper is organized as follows. Section II presents the signal model and problem formulation. Section III proposes the data fusion algorithm and simplify the solutions to reduce the computational complexity. In Section IV, semi-real dataset experiments are employed to evaluate the performance of the proposed algorithms. Then, some conclusions are drawn in Section V.

Notation: $\mathbb{R}$, $\mathbb{R}^n$ and $\mathbb{R}^{m\times n}$ denote the set of real number, n-vector and $m \times n$ matrices, respectively. $\mathbb{R}_+$, $\mathbb{R}^n_+$ and $\mathbb{R}^{m\times n}_+$ denote the set of nonnegative real number, n-vector and $m \times n$ matrices, respectively. $\|\cdot\|_p$ denotes the p-norm. $\|\cdot\|_F$ represents the Frobenius-norm. $\mathcal{I}_L \triangleq \{1, \ldots, L\}$ stands for any positive integer L. Boldface $\mathbf{0}$, $\mathbf{1}$ and $\mathbf{I}$ denote the all-zero vector, all-one vector and identity matrix of proper dimension, respectively. $\mathbf{e}_i^{(m)}$ denotes the ith m-dimensional unit vector. conv($\mathcal{S}$) denotes the convex hull of the set $\mathcal{S}$ [33]. $\mathbf{vec}(\mathbf{X})$ represents the vector formed by stacking the columns of the matrix $\mathbf{X}$. $\otimes$ stands for the Kronecker product. $[\cdot]_+$ denotes the orthogonal projection onto the nonnegative orthant of Euclidean space. $\succeq$ denotes the componentwise inequality operation.

II. SIGNAL MODEL AND PROBLEM FORMULATION

Data fusion aims at fusing the observed low-spatial-resolution HSI data $\mathbf{Y}_h \in \mathbb{R}^{M\times L_h}$ and high-spatial-resolution MSI data $\mathbf{Y}_m \in \mathbb{R}^{M_m\times L}$ to yield the desired hyperspectral data $\mathbf{Z} \in \mathbb{R}^{M\times L}$, where M and L_h (or M_m and L) represent the number of spectral bands and the number of pixels in $\mathbf{Y}_h$ (or $\mathbf{Y}_m$), respectively. In other words, $\mathbf{Y}_h$ and $\mathbf{Y}_m$ can

be obtained by degrading the desired hyperspectral data $\mathbf{Z}$ spatially (i.e., $L_h < L$) and spectrally (i.e., $M_m < M$), respectively.

Let N denote the number of endmembers in the selected region. The high-spatial-resolution hyperspectral data $\mathbf{Z}$ can be decomposed into two matrices in hyperspectral unmixing by the linear mixing model [23] as

$$\mathbf{Z} = \mathbf{AS}, \tag{1}$$

where $\mathbf{A} \succeq \mathbf{0}_{M\times N}$ is the endmember signature matrix with ith column being the ith endmember, and $\mathbf{S} \succeq \mathbf{0}_{N\times L}$ is the abundance matrix [2], [34]. To be exact, the observed mode [23] is expressed as

$$\mathbf{Y}_h = \mathbf{ZG} + \mathbf{E}_h = \mathbf{ASG} + \mathbf{E}_h, \tag{2}$$

$$\mathbf{Y}_m = \mathbf{FZ} + \mathbf{E}_m = \mathbf{FAS} + \mathbf{E}_m, \tag{3}$$

where $\mathbf{E}_h$ and $\mathbf{E}_m$ are the residuals. The spectral response transform matrix $\mathbf{F} \in \mathbb{R}^{M_m\times M}$ downsamples the hyperspectral bands of $\mathbf{Z}$, yielding the multispectral data $\mathbf{Y}_m$. And the spatial spread transform matrix

$$\mathbf{G} = \mathbf{I}_{L_h} \otimes \mathbf{g} \in \mathbb{R}^{L\times L_h}, \tag{4}$$

demonstrated in Fig. 1, blurs and downsamples the desired data $\mathbf{Z}$ to yield the low-spatial-resolution data $\mathbf{Y}_h$, in which $\mathbf{g} \in \mathbb{R}^{r^2}$ is associated with the Gaussian point spread function with a blurring factor of $r \triangleq \sqrt{L/L_h}$ [23], [35]. $\mathbf{F}$ and $\mathbf{G}$ are assumed to be known in this article [23], [31].

To address the ill-posedness of the original CNMF problem, total variation and signature-based regularizers are introduced into this problem. By the equations (1), (2) and (3), the regularized CNMF criterion reconstructs the fused data $\mathbf{Z} = \mathbf{AS}$ by solving the following problem

$$\begin{aligned} &\min_{\mathbf{A},\mathbf{S}} \frac{1}{2} C(\mathbf{A},\mathbf{S}) + \lambda_a \phi_a(\mathbf{A}) + \lambda_s \phi_s(\mathbf{S}) \\ &\text{s.t. } \mathbf{A} \succeq 0_{M\times N}, \quad \mathbf{S} \succeq 0_{N\times L}, \end{aligned} \tag{5}$$

where $C(\mathbf{A},\mathbf{S}) \triangleq \|\mathbf{Y}_h - \mathbf{ASG}\|_F^2 + \|\mathbf{Y}_m - F\mathbf{AS}\|_F^2$ is the original CNMF criterion proposed in [23]; $\lambda_a > 0$ and $\lambda_s > 0$ are the regularization parameters.

In the problem (5), the regularizers $\phi_a(\mathbf{A})$ and $\phi_s(\mathbf{S})$ based on the volume of the simplex $\text{conv}\{\mathbf{a}_1,\ldots,\mathbf{a}_N\}$ are assigned to regulate the signature and abundance matrices, respectively. In this article, two regularization expressions are introduced into the original CNMF as follows [33], [36],

$$\phi_a(\mathbf{A}) \triangleq \frac{1}{2}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \|\mathbf{a}_i - \mathbf{a}_j\|_2^2, \tag{6a}$$

$$\phi_s(\mathbf{S}) \triangleq \sum_{\{i,j\}\in\epsilon} \|\mathbf{s}_i - \mathbf{s}_j\|_1, \tag{6b}$$

where ϵ denotes the set of horizontal and vertical neighbors in the image. $\phi_a(\mathbf{A})$ is the well-known iterated-constrained-endmembers (ICE) regularizer in the hyperspectral unmixing context [31], [36], while $\phi_s(\mathbf{S})$ is the anisotropic total variation regularizer that promotes piecewise smoothing in the fractional abundance of the same endmember among neighboring pixels [22], [37].

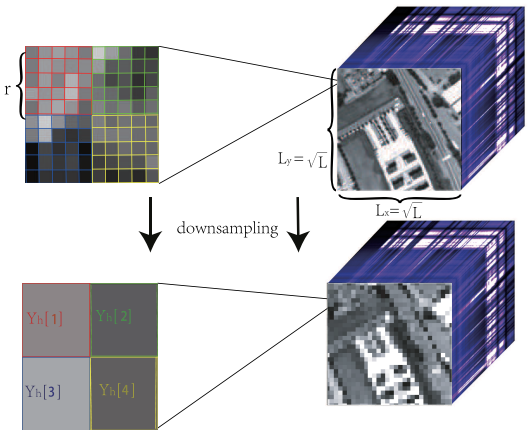

FIGURE 1. Spatial blurring effect by the matrix G with the parameter $\mathbf{g} \in \mathbb{R}^{25}$ ($r = 5$) for yielding the degraded image.

III. ADMM-BASED DATA FUSION ALGORITHM

In this section, we propose a novel CNMF algorithm with total variation and signature-based regularizers (termed as TVSR-CNMF) for efficiently solving the problem (5).

A. CONVEX REFORMULATION

Since (5) is a nonconvex problem, then it should be reformulated as a convex form. First of all, (6a) can be rewritten via convex optimization method [31], [33], [38] as

$$\phi(\mathbf{A}) = \frac{1}{2}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \|\mathbf{P}_{ij}\mathbf{a}\|_2^2, \tag{7}$$

where $\mathbf{a} \triangleq \mathbf{vec}(\mathbf{A})$, $\mathbf{P}_{ij} \triangleq (\mathbf{e}_i^{(N)} - \mathbf{e}_j^{(N)})^T \otimes \mathbf{I}_M$ and $\mathbf{P} \in \mathbb{R}^{0.5MN(N-1)\times MN}$. It can be easily proved that (5) is convex in $\mathbf{A}$ for each fixed $\mathbf{S}$. Then, we focus on the reformulation of (6b) from the nonconvex expression into the convex.

Assume that $\bar{\mathbf{S}}$ is an $L_y \times L_x$ square matrix (i.e. $L_y = L_x = \sqrt{L}$) for a single band of hyperspectral data cube $\mathbf{Z}$, depicted in Fig. 1, which is used to generate the horizontal and vertical difference matrices [37], [39], respectively. The vertical differences of $\bar{\mathbf{S}}$ can be presented as $\bar{\mathbf{R}}_{L_y}\bar{\mathbf{S}}$, where $\bar{\mathbf{R}}_{L_y}$ is the first-order vertical difference matrix [33], defined as

$$\bar{\mathbf{R}}_{L_y}(i,j) \triangleq \begin{cases} 1, & j = i+1, \\ -1, & j = i, \\ 0, & \text{otherwise}. \end{cases}$$

The vertical differences can also be computed via the vector-matrix operator by $\mathbf{vec}(\bar{\mathbf{R}}_{L_y}\bar{\mathbf{S}}) = \bar{\mathbf{D}}_y\bar{\mathbf{s}}$, where $\bar{\mathbf{D}}_y = \mathbf{I}_{L_x} \otimes \bar{\mathbf{R}}_{L_y}$ and $\bar{\mathbf{s}} = \mathbf{vec}(\bar{\mathbf{S}})$. Since there are N endmembers in each pixel as mentioned above, then the vectorized vertical difference matrix $\mathbf{D}_v \in \mathbb{R}^{N\sqrt{L}(\sqrt{L}-1)\times NL}$ for the abundance vector $\mathbf{s}$ is presented as

$$\mathbf{D}_v = \mathbf{I}_{L_x} \otimes \bar{\mathbf{R}}_{L_y} \otimes \mathbf{I}_N. \tag{8}$$

Algorithm 1 TVSR-CNMF Algorithm for Solving (5)

1: **Input:** $\mathbf{Y}_h$, $\mathbf{Y}_m$, $\mathbf{F}$ and $\mathbf{G}$.
2: **Initialize** $\mathbf{A}^0$.
3: $k = 0$
4: **while** the stopping criterion is not met **do**
5: compute $\mathbf{S}^{k+1}$ by (11);
6: compute $\mathbf{A}^{k+1}$ by (12);
7: $k := k + 1$;
8: **end while**
9: **Output** $\mathbf{Z} = \mathbf{A}^k\mathbf{S}^k$.

In the same way, the horizontal differences $\bar{\mathbf{S}}\bar{\mathbf{R}}_{L_x}^T$ can be vectorized into the expression $\bar{\mathbf{D}}_x\bar{\mathbf{s}}$, where $\bar{\mathbf{D}}_x = \bar{\mathbf{R}}_{L_x} \otimes I_{L_y}$ and $\bar{\mathbf{R}}_{L_x} = \bar{\mathbf{R}}_{L_y}$. Thus, the vectorized horizontal difference matrix can be computed as

$$\mathbf{D}_h = \bar{\mathbf{R}}_{L_x} \otimes I_{L_y} \otimes \mathbf{I}_N. \tag{9}$$

To sum up, (6b) can be reformulated via two convex surrogates as

$$\phi_s(\mathbf{S}) = \|\mathbf{D}_v\mathbf{s}\|_1 + \|\mathbf{D}_h\mathbf{s}\|_1, \tag{10}$$

where $\mathbf{s} \triangleq \mathbf{vec}(\mathbf{S}) \in \mathbb{R}^{NL}$. Different from the method in [22] that combined two difference matrices together, (10) handles two differences separately to implement variable splitting easily for complexity reduction. So far, one can see that the problem (5) is also convex in $\mathbf{S}$ for each fixed $\mathbf{A}$. In other words, the problem (5) is a bi-convex problem, so alternating optimization (AO) method is adopted to get the closed-form solutions.

B. TVSR-CNMF ALGORITHM VIA AO

The TVSR-CNMF algorithm is proposed via AO, shown in Algorithm 1, to alternatively iterate the following two convex subproblems until convergence,

$$\mathbf{S}^{k+1} \in \arg\min_{\mathbf{S}\geq 0_{N\times L}} \frac{1}{2}C(\mathbf{A}^k, \mathbf{S}) + \lambda_v\|\mathbf{D}_v\mathbf{s}\|_1 + \lambda_h\|\mathbf{D}_h\mathbf{s}\|_1, \tag{11}$$

$$\mathbf{A}^{k+1} \in \arg\min_{\mathbf{A}\geq 0_{M\times N}} \frac{1}{2}C(\mathbf{A}, \mathbf{S}^k) + \frac{\lambda_a}{2}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\|\mathbf{P}_{ij}\mathbf{a}\|_2^2, \tag{12}$$

where k denotes the iteration number of AO method (also called outer iteration); λ_v and λ_h (corresponding to λ_s) are the regularization parameters of vertical and horizontal total variation, respectively. As shown in the problem (11), the 1st iteration of $\mathbf{S}$ (i.e. $k = 0$) needs an initial value $\mathbf{A}^0$, which can be gained by the successive projection algorithm (SPA) [34], [40]. Due to the large size of the problems (11) and (12), two ADMM-based algorithms are proposed to solve these convex subproblems efficiently in the following subsections.

C. SOLVING ABUNDANCE MATRIX VIA ADMM

Using ADMM [41] and vector-matrix operators, we combine two Frobenius-norm terms into an ℓ_2-norm term by vectorizing the matrix. The bottleneck for solving (11) lies in the heavy computations of $\mathbf{B}_1$, $\mathbf{D}_v$ and $\mathbf{D}_h$. As far as ADMM is concerned, there's more than one way to reformulate this problem into the ADMM form. To handle the bottleneck efficiently, we rewrite (11) in this form wherein the primal variables can be split into several separable blocks by introducing different equality constraints. That is, the subproblem (11) can be reformulated as

$$\begin{aligned}
&\min_{\mathbf{s},\mathbf{u},\mathbf{z},\mathbf{x},\mathbf{v}_1,\mathbf{v}_2} \frac{1}{2}\|\mathbf{B}_1\mathbf{s} - \mathbf{y}\|_2^2 + \lambda_v\|\mathbf{v}_1\|_1 + \lambda_h\|\mathbf{v}_2\|_1 + I_+(\mathbf{x})\\
&\text{s.t. } \mathbf{s} = \mathbf{u},\\
&\quad \mathbf{v}_1 = \mathbf{D}_v\mathbf{u},\\
&\quad \mathbf{s} = \mathbf{z},\\
&\quad \mathbf{v}_2 = \mathbf{D}_h\mathbf{z},\\
&\quad \mathbf{s} = \mathbf{x},
\end{aligned} \tag{13}$$

where $\mathbf{B}_1 \triangleq [(\mathbf{G}^T \otimes \mathbf{A}^k)^T, (\mathbf{I}_L \otimes \mathbf{F}\mathbf{A}^k)^T]^T \in \mathbb{R}^{(ML_h+LM_m)\times NL}$, $\mathbf{y} \triangleq [\mathbf{vec}(\mathbf{Y}_h)^T, \mathbf{vec}(\mathbf{Y}_m)^T]^T \in \mathbb{R}^{ML_h+LM_m}$, and the indicator function $I_+(\mathbf{x})$ is defined as

$$I_+(\mathbf{x}) \triangleq \begin{cases} 0, & \text{if } \mathbf{x} \succeq \mathbf{0}_{MN},\\ \infty, & \text{otherwise.}\end{cases}$$

The augmented Lagrangian of (13) is presented by

$$\begin{aligned}
\mathcal{L}(\mathbf{s},\mathbf{u},\mathbf{z},\mathbf{x},\mathbf{v}_1,\mathbf{v}_2) = {} & \frac{1}{2}\|\mathbf{B}_1\mathbf{s} - \mathbf{y}\|_2^2 + \lambda_v\|\mathbf{v}_1\|_1 + \lambda_h\|\mathbf{v}_2\|_1\\
& + I_+(\mathbf{x}) + \boldsymbol{h}_1{}^T(\mathbf{s} - \mathbf{u}) + \frac{\eta}{2}\|\mathbf{s} - \mathbf{u}\|_2^2\\
& + \boldsymbol{h}_2{}^T(\mathbf{v}_1 - \mathbf{D}_v\mathbf{u}) + \frac{\eta}{2}\|\mathbf{v}_1 - \mathbf{D}_v\mathbf{u}\|_2^2\\
& + \boldsymbol{h}_3{}^T(\mathbf{s} - \mathbf{z}) + \frac{\eta}{2}\|\mathbf{s} - \mathbf{z}\|_2^2\\
& + \boldsymbol{h}_4{}^T(\mathbf{v}_2 - \mathbf{D}_h\mathbf{z}) + \frac{\eta}{2}\|\mathbf{v}_2 - \mathbf{D}_h\mathbf{z}\|_2^2\\
& + \boldsymbol{h}_5{}^T(\mathbf{s} - \mathbf{x}) + \frac{\eta}{2}\|\mathbf{s} - \mathbf{x}\|_2^2,
\end{aligned} \tag{14}$$

where $\boldsymbol{h}_i (i = 1, \cdots, 5)$ are dual variables, and η is the augmented Lagrangian parameter. Then, ADMM iteratively updates the primal and the dual variables in step by

$$\mathbf{s}^{j+1} \in \arg\min_{\mathbf{s}\in\mathbb{R}^{NL}} \mathcal{L}(\mathbf{s}, \mathbf{u}^j, \mathbf{z}^j, \mathbf{x}^j, \boldsymbol{h}_1{}^j, \boldsymbol{h}_3{}^j, \boldsymbol{h}_5{}^j), \tag{15a}$$

$$\mathbf{u}^{j+1} \in \arg\min_{\mathbf{u}\in\mathbb{R}^{NL}} \mathcal{L}(\mathbf{s}^{j+1}, \mathbf{u}, \mathbf{v}_1^j, \boldsymbol{h}_1{}^j, \boldsymbol{h}_2{}^j), \tag{15b}$$

$$\mathbf{v}_1^{j+1} \in \arg\min_{\mathbf{v}_1\in\mathbb{R}^{N\sqrt{L}(\sqrt{L}-1)}} \mathcal{L}(\mathbf{u}^{j+1}, \mathbf{v}_1, \boldsymbol{h}_2{}^j), \tag{15c}$$

$$\mathbf{z}^{j+1} \in \arg\min_{\mathbf{z}\in\mathbb{R}^{NL}} \mathcal{L}(\mathbf{s}^{j+1}, \mathbf{z}, \mathbf{v}_2^j, \boldsymbol{h}_3{}^j, \boldsymbol{h}_4{}^j), \tag{15d}$$

$$\mathbf{v}_2^{j+1} \in \arg\min_{\mathbf{v}_2\in\mathbb{R}^{N\sqrt{L}(\sqrt{L}-1)}} \mathcal{L}(\mathbf{z}^{j+1}, \mathbf{v}_2, \boldsymbol{h}_4{}^j), \tag{15e}$$

$$\mathbf{x}^{j+1} \in \arg\min_{\mathbf{x}\in\mathbb{R}^{NL}} \mathcal{L}(\mathbf{s}^{j+1}, \mathbf{x}, \boldsymbol{h}_5{}^j), \tag{15f}$$

$$\boldsymbol{h}_1{}^{j+1} = \boldsymbol{h}_1{}^j + \eta(\mathbf{s}^{j+1} - \mathbf{u}^{j+1}), \tag{15g}$$

$$\boldsymbol{h}_2{}^{j+1} = \boldsymbol{h}_2{}^j + \eta(\mathbf{v}_1^{j+1} - \mathbf{D}_v\mathbf{u}^{j+1}), \tag{15h}$$

$$\boldsymbol{h}_3^{j+1} = \boldsymbol{h}_3^{j} + \eta(\mathbf{s}^{j+1} - \mathbf{z}^{j+1}), \tag{15i}$$
$$\boldsymbol{h}_4^{j+1} = \boldsymbol{h}_4^{j} + \eta(\mathbf{v}_2^{j+1} - \mathbf{D}_h\mathbf{z}^{j+1}), \tag{15j}$$
$$\boldsymbol{h}_5^{j+1} = \boldsymbol{h}_5^{j} + \eta(\mathbf{s}^{j+1} - \mathbf{x}^{j+1}), \tag{15k}$$

where j stands for the iteration number of ADMM-based algorithms (also called inner iteration). Furthermore, (15a) $\cdots$ (15f) can be conveniently given in the following scaled forms as

$$\mathbf{s}^{j+1} \in \arg\min_{\mathbf{s}\in\mathbb{R}^{NL}} \frac{1}{2}\|\mathbf{B}_1\mathbf{s}-\mathbf{y}\|_2^2 + \frac{\eta}{2}\|\mathbf{s}-\mathbf{u}^j+\widetilde{\boldsymbol{h}}_1^j\|_2^2 + \frac{\eta}{2}\|\mathbf{s}-\mathbf{z}^j+\widetilde{\boldsymbol{h}}_3^j\|_2^2 + \frac{\eta}{2}\|\mathbf{s}-\mathbf{x}^j+\widetilde{\boldsymbol{h}}_5^j\|_2^2, \tag{16a}$$

$$\mathbf{u}^{j+1} \in \arg\min_{\mathbf{u}\in\mathbb{R}^{NL}} \frac{\eta}{2}\|\mathbf{s}^{j+1}-\mathbf{u}+\widetilde{\boldsymbol{h}}_1^j\|_2^2 + \frac{\eta}{2}\|\mathbf{v}_1^j-\mathbf{D}_v\mathbf{u}+\widetilde{\boldsymbol{h}}_2^j\|_2^2, \tag{16b}$$

$$\mathbf{v}_1^{j+1} \in \arg\min_{\mathbf{v}_1\in\mathbb{R}^{N\sqrt{L}(\sqrt{L}-1)}} \lambda_v\|\mathbf{v}_1\|_1 + \frac{\eta}{2}\|\mathbf{v}_1-\mathbf{D}_v\mathbf{u}^{j+1}+\widetilde{\boldsymbol{h}}_2^j\|_2^2, \tag{16c}$$

$$\mathbf{z}^{j+1} \in \arg\min_{\mathbf{z}\in\mathbb{R}^{NL}} \frac{\eta}{2}\|\mathbf{s}^{j+1}-\mathbf{z}+\widetilde{\boldsymbol{h}}_3^j\|_2^2 + \frac{\eta}{2}\|\mathbf{v}_2^j-\mathbf{D}_h\mathbf{z}+\widetilde{\boldsymbol{h}}_4^j\|_2^2, \tag{16d}$$

$$\mathbf{v}_2^{j+1} \in \arg\min_{\mathbf{v}_2\in\mathbb{R}^{N\sqrt{L}(\sqrt{L}-1)}} \lambda_h\|\mathbf{v}_2\|_1 + \frac{\eta}{2}\|\mathbf{v}_2-\mathbf{D}_h\mathbf{u}^{j+1}+\widetilde{\boldsymbol{h}}_2^j\|_2^2, \tag{16e}$$

$$\mathbf{x}^{j+1} \in \arg\min_{\mathbf{x}\in\mathbb{R}^{NL}} I_+(\mathbf{x}) + \frac{\eta}{2}\|\mathbf{s}^{j+1}-\mathbf{x}+\widetilde{\boldsymbol{h}}_5^j\|_2^2, \tag{16f}$$

where $\widetilde{\boldsymbol{h}}_i \triangleq \boldsymbol{h}_i/\eta, (i = 1, \cdots, 5)$ are defined as the scaled dual variables; all the variables, including $\widetilde{\boldsymbol{h}}_i^0$, $\mathbf{u}^0$, $\mathbf{z}^0$, $\mathbf{x}^0$, $\mathbf{v}_1^0$ and $\mathbf{v}_2^0$, are initialized by zero vectors with proper sizes or warm start [41].

By observational analysis, (16a), (16b) and (16d) are all unconstrained quadratic problems, (16c) and (16e) are the generalized Lasso problems [41], and (16f) is a proximal operator with $I_+(\mathbf{x})$. Thus, the closed-form solutions are derived and simplified to reduce the complexity step by step. The closed-form solution of (16a) is expressed as

$$\mathbf{s}^{j+1} = (\mathbf{B}_1^T\mathbf{B}_1 + 3\eta\mathbf{I}_{NL})^{-1}[\mathbf{B}_1^T\mathbf{y} + \eta(\mathbf{u}^j - \widetilde{\boldsymbol{h}}_1^j + \mathbf{z}^j - \widetilde{\boldsymbol{h}}_3^j + \mathbf{x}^j - \widetilde{\boldsymbol{h}}_5^j)]. \tag{17}$$

The computational complexity of (17) is shown as $\mathcal{O}\left(L_h(NL)^2\xi\right)$, where $\xi \triangleq \max\{M, M_m r^2, Nr^2\}$. It is obvious that the complexity of $\mathbf{B}_1^T\mathbf{B}_1$ is the bottleneck to increase the efficiency of ADMM iterations, especially when the size of L reaches tens of thousands. To conduct the complexity reduction, we define $\bar{\mathbf{B}}_1 \triangleq [(\mathbf{g}^T \otimes \mathbf{A}^k)^T, (\mathbf{I}_{r^2} \otimes \mathbf{F}\mathbf{A}^k)^T]^T \in \mathbb{R}^{(M+r^2M_m)\times Nr^2}$, $\mathbf{B}_1^T\mathbf{y} \triangleq \mathbf{vec}((\mathbf{g}^T \otimes \mathbf{A}^k)^T\mathbf{Y}_h) + \mathbf{vec}((\mathbf{F}\mathbf{A}^k)^T\mathbf{Y}_m)$, $\tilde{\mathbf{v}} \triangleq \mathbf{B}_1^T\mathbf{y} + \eta(\mathbf{u}^j - \widetilde{\boldsymbol{h}}_1^j + \mathbf{z}^j - \widetilde{\boldsymbol{h}}_3^j + \mathbf{x}^j - \widetilde{\boldsymbol{h}}_5^j)$, and then reshape $\tilde{\mathbf{v}}$ into a matrix $\widetilde{\mathbf{V}} \in \mathcal{R}^{Nr^2\times L_h}$. Using structure feature of the matrix (4) and $\mathbf{I}_L = \mathbf{I}_{L_h} \otimes \mathbf{I}_{r^2}$, the optimized solution of (17) is rewritten as

$$\mathbf{s}^{j+1} = \mathbf{vec}((\bar{\mathbf{B}}_1^T\bar{\mathbf{B}}_1 + 3\eta\mathbf{I}_{Nr^2})^{-1}\widetilde{\mathbf{V}}). \tag{18}$$

The complexity of (18) is no more than $\mathcal{O}(((Nr^2)^2 + NL)\xi)$, whose complexity order is much lower than that of (17). More details are given in the Appendix of [31].

As for (16b), the closed-form solution is presented as

$$\mathbf{u}^{j+1} = (\mathbf{D}_v^T\mathbf{D}_v + \eta\mathbf{I}_{NL})^{-1}[\mathbf{D}_v^T(\mathbf{v}_1^j + \widetilde{\boldsymbol{h}}_2^j) + \mathbf{s}^{j+1} + \widetilde{\boldsymbol{h}}_1^j], \tag{19}$$

which amounts to a high complexity of $\mathcal{O}\left((NL)^3\right)$. Let $\tilde{\mathbf{u}}_1 \triangleq \mathbf{v}_1^j + \widetilde{\boldsymbol{h}}_2^j$ and $\tilde{\mathbf{u}}_2 \triangleq \mathbf{s}^{j+1} + \widetilde{\boldsymbol{h}}_1^j$, which are then reshaped into the matrices $\widetilde{\mathbf{U}}_1$ and $\widetilde{\mathbf{U}}_2$, respectively. By the structure of (8), (19) is simplified as

$$\mathbf{u}^{j+1} = \mathbf{vec}(\{[(\bar{\mathbf{R}}_{L_y}^T\bar{\mathbf{R}}_{L_y} + \eta\,\mathbf{I}_{L_y})^{-1}\bar{\mathbf{R}}_{L_y}^T] \otimes \mathbf{I}_N\}\widetilde{\mathbf{U}}_1) + \mathbf{vec}([(\bar{\mathbf{R}}_{L_y}^T\bar{\mathbf{R}}_{L_y} + \eta\,\mathbf{I}_{L_y})^{-1} \otimes \mathbf{I}_N]\widetilde{\mathbf{U}}_2) \tag{20}$$

which involves the computational complexity of $\mathcal{O}\left(N^2L\sqrt{L}\right)$, detailedly given in Appendix A. Based on these results above, (16c) has the following closed-form solution as

$$\mathbf{v}_1^{j+1} = \text{shrink}(\mathbf{D}_v\mathbf{u}^{j+1} - \widetilde{\boldsymbol{h}}_2^j, \lambda_v/\eta), \tag{21}$$

where the shrinkage operator is defined as

$$\text{shrink}(\mathbf{v}, \lambda) = \text{sgn}(\mathbf{v})\max(|\mathbf{v}| - \lambda, 0). \tag{22}$$

Similar to (16b), the closed-form solution of (16d) is given as

$$\mathbf{z}^{j+1} = (\mathbf{D}_h^T\mathbf{D}_h + \eta\mathbf{I}_{NL})^{-1}[\mathbf{D}_h^T(\mathbf{v}_2^j + \widetilde{\boldsymbol{h}}_4^j) + \mathbf{s}^{j+1} + \widetilde{\boldsymbol{h}}_3^j], \tag{23}$$

which has the same computational complexity of $\mathcal{O}\left((NL)^3\right)$ as (19). Let $\tilde{\mathbf{z}}_1 \triangleq \mathbf{v}_2^j + \widetilde{\boldsymbol{h}}_4^j$ and $\tilde{\mathbf{z}}_2 \triangleq \mathbf{s}^{j+1} + \widetilde{\boldsymbol{h}}_3^j$, which are then reshaped into the matrices $\widetilde{\mathbf{Z}}_1$ and $\widetilde{\mathbf{Z}}_2$, respectively. Thus, (23) can be simplified as

$$\mathbf{z}^{j+1} = \mathbf{vec}(\widetilde{\mathbf{Z}}_1\bar{\mathbf{R}}_{L_y}[(\bar{\mathbf{R}}_{L_y}^T\bar{\mathbf{R}}_{L_y} + \eta\,\mathbf{I}_{L_y})^{-1}]^T) + \mathbf{vec}(\widetilde{\mathbf{Z}}_2[(\bar{\mathbf{R}}_{L_y}^T\bar{\mathbf{R}}_{L_y} + \eta\,\mathbf{I}_{L_y})^{-1}]^T), \tag{24}$$

whose complexity is evaluated by $\mathcal{O}\left(NL\sqrt{L}\right)$, detailed in Appendix B. Correspondingly, we can get the closed-form solution of (16e) via shrinkage operator as

$$\mathbf{v}_2^{j+1} = \text{shrink}(\mathbf{D}_h\mathbf{z}^{j+1} - \widetilde{\boldsymbol{h}}_4^j, \lambda_h/\eta). \tag{25}$$

As we all know, (16f) is a projection onto the nonnegative orthant $\mathcal{R}_+^{NL}$ via the indicator function $I_+(\mathbf{x})$, yielding the following closed-form solution:

$$\mathbf{x}^{j+1} = \left[\mathbf{s}^{j+1} + \widetilde{\boldsymbol{h}}_5^j\right]_+. \tag{26}$$

All in all, the resulting ADMM algorithm is shown in Algorithm 2.

Algorithm 2 ADMM Algorithm for Solving (11)

Input: N, $\mathbf{Y}_h$, $\mathbf{Y}_m$, $\mathbf{F}$, $\mathbf{G}$, and $\mathbf{A}^k$.
Initialize $\mathbf{u}^0$, $\mathbf{z}^0$, $\mathbf{x}^0$,$\mathbf{v}_1^0$, $\mathbf{v}_2^0$ and $\widetilde{\boldsymbol{h}}_i^0 (i = 1, \ldots, 5)$ with $\mathbf{0}$ or warm start.
$j = 0$
while the stopping criterion is not met **do**
  compute $\mathbf{s}^{j+1}$ by (18);
  compute $\mathbf{u}^{j+1}$ by (20);
  compute $\mathbf{v}_1^{j+1}$ by (21);
  compute $\mathbf{z}^{j+1}$ by (24);
  compute $\mathbf{v}_2^{j+1}$ by (25);
  compute $\mathbf{x}^{j+1}$ by (26);
  compute $\widetilde{\boldsymbol{h}}_i^{j+1} = \boldsymbol{h}_i^{j+1}/\eta$, $(i = 1, \cdots, 5)$ by (15g) $\cdots$ (15k), respectively;
  $j := j + 1$;
end while.
Output $\mathbf{S}^{k+1}$.

D. SOLVING ENDMEMBER SIGNATURE MATRIX VIA ADMM

By observation, the bottleneck in (12) comes from the non-negativity constraint, which amounts to the projection of the solution onto the non-negative orthant. To solve (12) by ADMM [41], we rewrite (12) in a certain form such that the primal variable can be split into several parts. To be precise, we reformulate (12) as

$$\min_{\mathbf{a},\boldsymbol{\beta}\in\mathbb{R}^{MN}} \frac{1}{2}\|\mathbf{B}_2\mathbf{a}-\mathbf{y}\|_2^2 + \frac{\lambda_a}{2}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\|\mathbf{P}_{ij}\mathbf{a}\|_2^2 + I_+(\boldsymbol{\beta}) \quad \text{s.t. } \mathbf{a}=\boldsymbol{\beta} \tag{27}$$

where $\mathbf{B}_2 \triangleq [((\mathbf{S}^{k+1}\mathbf{G})^T \otimes \mathbf{I}_M)^T, ((\mathbf{S}^{k+1})^T \otimes \mathbf{F})^T]^T \in \mathbb{R}^{(ML_h+LM_m)\times MN}$ and $\mathbf{a} = \mathbf{vec}(\mathbf{A}) \in \mathbb{R}_+^{MN}$.

The augmented Lagrangian of the problem (27) is then given by

$$\mathcal{L}(\mathbf{a},\boldsymbol{\beta},\mathbf{h}) = \frac{1}{2}\|\mathbf{B}_2\mathbf{a}-\mathbf{y}\|_2^2 + \frac{\lambda_a}{2}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\|\mathbf{P}_{ij}\mathbf{a}\|_2^2 + I_+(\boldsymbol{\beta}) + \mathbf{h}^T(\mathbf{a}-\boldsymbol{\beta}) + \frac{\eta}{2}\|\mathbf{a}-\boldsymbol{\beta}\|_2^2,$$

where $\mathbf{h} \in \mathbb{R}^{MN}$ is the dual variable associated with the equality constraint in (27), and $\eta > 0$ is the penalty parameter. Then, ADMM iteratively updates the two primal variables and the dual variable, with provable convergence property [42], as follows:

$$\mathbf{a}^{j+1} \in \arg\min_{\mathbf{a}\in\mathbb{R}^{MN}} \mathcal{L}(\mathbf{a},\boldsymbol{\beta}^j,\mathbf{h}^j), \tag{28a}$$

$$\boldsymbol{\beta}^{j+1} \in \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^{MN}} \mathcal{L}(\mathbf{a}^{j+1},\boldsymbol{\beta},\mathbf{h}^j), \tag{28b}$$

$$\mathbf{h}^{j+1} = \mathbf{h}^j + \eta(\mathbf{a}^{j+1}-\boldsymbol{\beta}^{j+1}), \tag{28c}$$

where $\boldsymbol{\beta}^0$ and $\mathbf{h}^0$ are initialized by $\mathbf{0}_{MN}$ or warm start. (28a) can be further expressed as an unconstrained quadratic convex problem, that is,

$$\mathbf{a}^{j+1} \in \arg\min_{\mathbf{a}\in\mathbb{R}^{MN}} \frac{1}{2}\|\mathbf{B}_2\mathbf{a}-\mathbf{y}\|_2^2 + \frac{\lambda_a}{2}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\|\mathbf{P}_{ij}\mathbf{a}\|_2^2 + \mathbf{h}^{jT}(\mathbf{a}-\boldsymbol{\beta}^j) + \frac{\eta}{2}\|\mathbf{a}-\boldsymbol{\beta}^j\|_2^2. \tag{29}$$

Thus, we can derive the closed-form solution as

$$\mathbf{a}^{j+1} = (\mathbf{B}_2^T\mathbf{B}_2 + \lambda_a\mathbf{P}^T\mathbf{P} + \eta\mathbf{I}_{MN})^{-1}(\mathbf{B}_2^T\mathbf{y} + \eta\boldsymbol{\beta}^j - \mathbf{h}^j), \tag{30}$$

where $\mathbf{P} \in \mathbb{R}^{(0.5MN(N-1))\times(MN)}$ is the matrix formed by stacking all the $\mathbf{P}_{ij} \in \mathbb{R}^{M\times(MN)}$. The computational complexity of (30) is $\mathcal{O}(N^4M^3 + (NM)^2\xi')$, where $\xi' \triangleq \max\{ML_h, M_mL\}$. Using the structure of $\mathbf{B}_2$, (30) can be presented as

$$\mathbf{a}^{j+1} = \{((\mathbf{S}^{k+1}\mathbf{G})(\mathbf{S}^{k+1}\mathbf{G})^T)\otimes\mathbf{I}_M + \lambda_a\mathbf{P}^T\mathbf{P} + \eta\mathbf{I}_{MN} + (\mathbf{S}^{k+1}(\mathbf{S}^{k+1})^T)\otimes\mathbf{F}^T\mathbf{F}\}^{-1}(\mathbf{B}_2^T\mathbf{y} + \eta\boldsymbol{\beta}^j - \mathbf{h}^j) \tag{31}$$

where $\mathbf{B}_2^T\mathbf{y} = \mathbf{vec}(\mathbf{Y}_h(\mathbf{S}^{k+1}\mathbf{G})^T) + \mathbf{vec}(\mathbf{F}^T\mathbf{Y}_m(\mathbf{S}^{k+1})^T)$, detailedly given in the Appendix of [31]. The computation of (31) is evaluated by $\mathcal{O}\left(N^4M^3 + N^2L\right)$, which is much lower than that of (30).

By defining the scaled dual variable $\boldsymbol{\mu} \triangleq \mathbf{h}/\eta$ [41], (28b) can be conveniently expressed in the following scaled form as

$$\boldsymbol{\beta}^{j+1} \in \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^{MN}} I_+(\boldsymbol{\beta}) + \frac{\eta}{2}\|\mathbf{a}^{j+1} - \boldsymbol{\beta} + \boldsymbol{\mu}^j\|_2^2 \tag{32}$$

which is generally referred to as a proximity operator for the indicator function $I_+(\boldsymbol{\beta})$ [41], with a closed-form solution given by

$$\boldsymbol{\beta}^{j+1} = [\mathbf{a}^{j+1} + \boldsymbol{\mu}^j]_+. \tag{33}$$

In sum, the resulting ADMM algorithm is summarized in Algorithm 3.

Algorithm 3 ADMM Algorithm for Solving (12)

Input: $\mathbf{Y}_h$, $\mathbf{Y}_m$, $\mathbf{B}$, $\mathbf{D}$, and $\mathbf{S}^{k+1}$.
Initialize $\boldsymbol{\beta}^0$ and $\mathbf{h}^0$ with $\mathbf{0}_{MN}$ or warm start.
$j = 0$
while the stopping criterion is not met **do**
  compute $\mathbf{a}^{j+1}$ by (31);
  compute $\boldsymbol{\beta}^{j+1}$ by (32);
  compute $\boldsymbol{\mu}^{j+1} = \mathbf{h}^{j+1}/\eta$ by (28c);
  $j := j + 1$;
end while
Output $\mathbf{A}^{k+1}$.

IV. EXPERIMENTS AND PERFORMANCE ANALYSIS

This section conducts the experimental tests to evaluate the performance of the proposed TVSR-CNMF algorithm. Wald's protocol is adopted to design the experiments [7], [23], [32], in which the observed hyperspectral datasets are regarded as the reference images $\mathbf{Z}$

to simulate the degraded images $\mathbf{Y}_h$ and $\mathbf{Y}_m$, yielding the fused image $\widehat{\mathbf{Z}}$. To measure the similarity quantitatively between $\widehat{\mathbf{Z}}$ and $\mathbf{Z}$, a lot of widely used quality metrics are utilized [7], [23], [25], [31], including reconstructed signal-to-noise ratio (RSNR), root mean squared error (RMSE), spectral angle mapper (SAM), and *erreur relative globale adimensionnelle de synthèse* (ERGAS). In addition, the computational time T (in seconds) is regarded as the measure of computational efficiency.

In order to reduce the effect of randomness, the performance in the experiments takes the corresponding mean of multiple measurements. In comparison, we take four state-of-the-art fusion methods, including convex optimization-based CNMF(CO-CNMF) [31], Bayesian method [25], Hysure [22] and the original CNMF [23] as the baseline algorithms. All the algorithms under test are implemented using Mathworks Matlab R2015a, running on a computer equipped with Xeon E5-2630 CPU with 2.40GHz speed and 64GB random access memory.

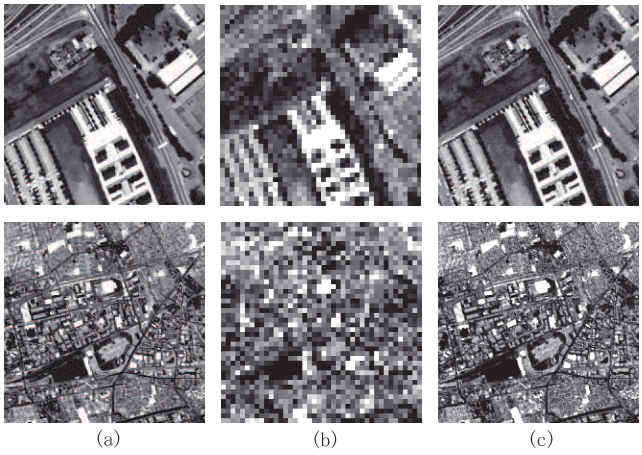

FIGURE 2. The 50th band of (a) the reference image Z, (b) the blurred low-spatial-resolution image $\mathbf{Y}_h$ and (c) the reconstructed image $\widehat{\mathbf{Z}}$ obtained by the proposed TVSR-CNMF algorithm for (top row) Pavia University and (bottom row) Moffett datasets, respectively.

A. DATASETS

In this paper, we consider two datasets that were acquired by two different hyperspectral sensors, respectively. The first dataset was taken over the area of Pavia University, northern Italy [43], by the reflective optics system imaging spectrometer (ROSIS) sensor, with 115 spectral bands and a spatial resolution of 1.3 m. After preprocessing, a total of 103 bands (ranging from 430 to 860 nm) are used in our experiments. The IKONOS [44] sensor that is utilized for the ROSIS sensor scene, approximates to Landsat TM bands 1-4 (covering 445-516-, 516-595, 632-698, and 757-853 nm regions). Thus, a spectral response transform matrix $\mathbf{F} \in \mathbb{R}^{4\times 103}$ uniformly downsamples the reference data $\mathbf{Z}$ to generate the multispectral data $\mathbf{Y}_m$ [22], [25], [31]. The second dataset was acquired by Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over Moffett Field, CA, in 1997 [45]. A total of 183 bands approximately correspond to the Landsat TM bands 1-5 and 7 after preprocessing, covering from 400 to 2500 nm. In consequence, a spectral response matrix $\mathbf{F} \in \mathbb{R}^{6\times 183}$ is constructed to yield the multispectral data $\mathbf{Y}_m$ for Moffett dataset [22], [23], [31], [46].

For these two reference images, both $L = 210 \times 210$ subscenes are selected as regions of interest [25], [31], [34], the 50th bands of which are displayed in the left column of Fig. 2. Thus, with the parameters setting of variance 2 and blurring factor $r = 5$ [23], [35], [47], a spatial spread transform matrix $\mathbf{G} \in \mathbb{R}^{44100\times 1764}$ was employed to downsample the reference data $\mathbf{Z}$ spatially to obtain the degraded hyperspectral data $\mathbf{Y}_h$ with low spatial resolution [35], [48], [49]. The associated 50th band image of $\mathbf{Y}_h$, with $L_h = 42 \times 42$ pixels, are displayed in the middle column of Fig. 2.

B. PARAMETER SETTING AND NUMBER OF ENDMEMBER

The optimized problem is reformulated into the regularized CNMF with three regularizers, as well as three corresponding weights λ_a, λ_v and λ_h, respectively. The values of these three weights are set as $\lambda_a = \lambda_v = \lambda_h = 0.001$ to balance the original CNMF terms and regularization terms [31], [47], which are relevant to the noise powers of HSIs and MSIs. The stopping rule is adopted that the relative difference between the successive updates of the objective function is less than 10^{-3}. Experiments have demonstrated that Algorithm 2 can converge within 5 iterations faster than Algorithm 1, and varying these parameters have no much impact on the convergence of the whole algorithm without running ADMM exhaustively [47]. Thus, the numbers of outer and inner iterations are set to be 30 and 10 for AO and ADMM algorithms [31], [47], respectively.

To study the denoising performance of the algorithm with total variation regularization, the SNR values of HSIs and MSIs range separately from 30 dB to 20 dB with an interval of 5 dB in the discussion. Fig. 3 shows that the performance (i.e. RSNR and RMSE) vary with the increase of noise (i.e. $\sigma^2/\mathrm{P}_{signal}$) for two datasets. We can observe that the performance of CO-CNMF algorithm is better than those of the proposed TVSR-CNMF and other baseline methods when the SNRs of $\mathbf{Y}_h$ and $\mathbf{Y}_m$ are equal or greater than about 30dB. It verifies that CO-CNMF did improve the performance greater than other methods under the condition of high SNRs, as described in [31]. However, with the decrease of SNRs, the performance of CO-CNMF algorithm degrades rapidly while our proposed TVSR-CNMF outperforms CO-CNMF and other baseline methods in Fig. 3, owing to the effect of total variation regularization.

The number of endmembers can be determined by widely used virtual dimensionality (VD) for two datasets [50], [51]. However, following the CNMF criterion [23], the number of endmember N can be set to be larger than the ground-truth value, due to shade and mixed endmembers in the regions [23]. To evaluate a proper value of N, we test the fusion performance of the proposed TVSR-CNMF algorithm w.r.t. different N for two datasets, demonstrated in Fig. 4.

TABLE 1. Performance comparison of fusion algorithms for Pavia University dataset.

Methods	Case I : SNR(Ym)=30, SNR(Yh)=25				Case II : SNR(Ym)=25, SNR(Yh)=20				Case III : SNR(Ym)=20, SNR(Yh)=20				Time T
	RSNR	RMSE	SAM	ERGAS	RSNR	RMSE	SAM	ERGAS	RSNR	RMSE	SAM	ERGAS	(secs)
CO-CNMF [31]	**25.623**	**90.238**	3.0408	**1.4026**	22.353	131.485	4.1789	1.8015	18.980	193.871	6.0663	2.6711	11.939
CNMF [23]	22.989	122.202	3.9897	1.7256	18.929	195.020	6.7348	2.5257	16.569	255.896	8.3162	3.2459	**11.484**
Bayesian [25]	21.924	138.143	3.4267	2.0696	21.251	149.278	3.8403	2.2082	19.845	175.498	**4.6259**	2.5804	197.539
Hysure [22]	20.815	156.965	4.4525	2.2685	20.038	171.637	5.4604	2.5773	17.884	219.960	6.2622	2.9430	27.795
Proposed	25.597	90.511	**3.1165**	1.4335	**23.262**	**118.428**	**3.6650**	**1.661**	**19.927**	**173.855**	5.0759	**2.3388**	43.269

TABLE 2. Performance comparison of fusion algorithms for Moffett dataset.

Methods	Case I : SNR(Ym)=30, SNR(Yh)=25				Case II : SNR(Ym)=25, SNR(Yh)=20				Case III : SNR(Ym)=20, SNR(Yh)=20				Time T
	RSNR	RMSE	SAM	ERGAS	RSNR	RMSE	SAM	ERGAS	RSNR	RMSE	SAM	ERGAS	(secs)
CO-CNMF [31]	**27.425**	**86.069**	2.139	**0.936**	23.789	130.819	3.348	1.379	19.700	209.467	5.202	2.195	20.248
CNMF [23]	23.991	127.808	2.820	1.372	20.522	190.550	4.761	1.959	17.907	257.490	6.557	2.646	**14.043**
Bayesian [25]	23.525	134.854	2.249	1.478	22.711	148.097	2.574	1.616	21.247	175.279	**3.190**	1.900	194.720
Hysure [22]	22.053	159.756	3.431	1.638	20.960	181.181	3.980	1.868	18.793	232.511	5.107	2.407	30.915
Proposed	27.256	87.761	**2.025**	0.953	**24.656**	**125.189**	**2.726**	**1.249**	**21.415**	**171.937**	4.019	**1.864**	54.952

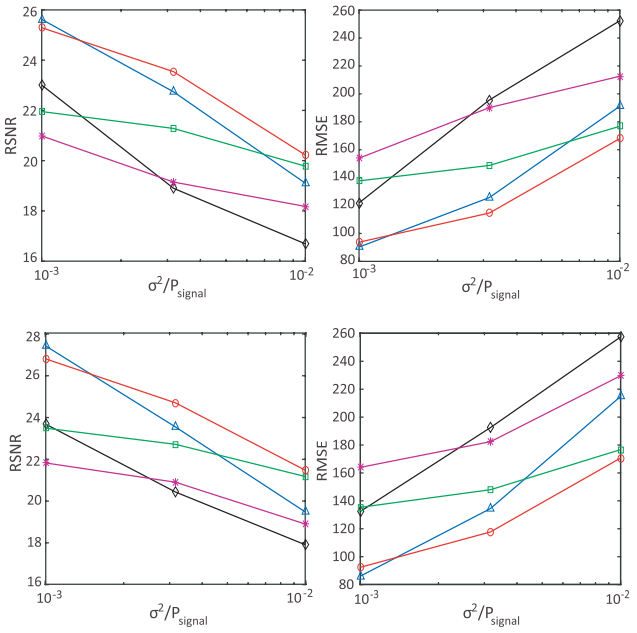

FIGURE 3. The curves of CO-CNMF (blue line with △), CNMF (black line with ⋄), Bayesian (green line with □), HySure (purple line with ∗) and the proposed TVSR-CNMF (red line with ○) algorithms, in terms of (left column) RSNR and (right column) RMSE, w.r.t. different the values of N, for (top row) Pavia University and (bottom row) Moffett datasets, respectively.

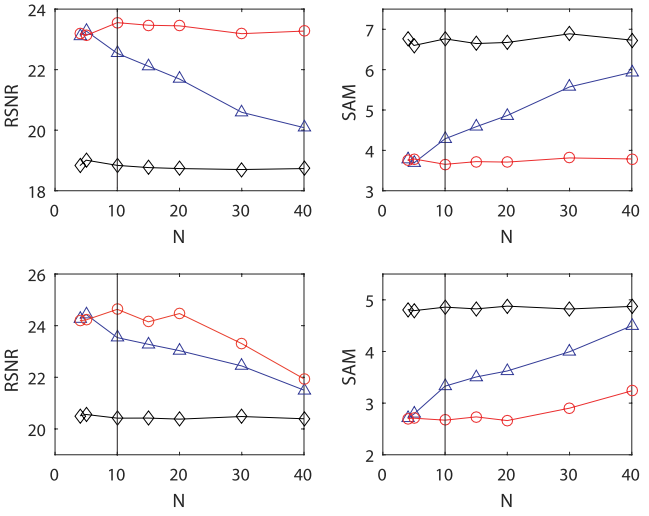

FIGURE 4. The curves of CO-CNMF (blue line with △), CNMF (black line with ⋄) and the proposed TVSR-CNMF (red line with ○) algorithms, in terms of (left column) RSNR and (right column) SAM w.r.t. different the values of N, for (top row) Pavia University and (bottom row) Moffett datasets, respectively.

From the figure, one can see that the performance curve varies with the value N to a certain extent. Also, setting the number of endmember $N > 10$ does not improve the data fusion performance too much (in terms of both RSNR and SAM), and especially the proposed TVSR-CNMF algorithm almost outperforms the other two CNMF-based algorithms for two datasets. Without loss of generality, we set $N = 10$ for the proposed TVSR-CNMF algorithm in all the subsequent experiments [31]. The reconstructed high-spatial-resolution images, obtained by the proposed TVSR-CNMF under the parameter setting above, are displayed in the right column of Fig. 2, where the reconstructed images hold high resemblance to their respective reference images. The performance differences between fusion algorithms are distinguished detailedly by the ensuing tables and curves, although the differences may be indiscernible visually.

C. PERFORMANCE COMPARISON AND DISCUSSION

The performance results of the proposed algorithm and four baseline methods are summarized in Table 1 and 2 for Pavia University and Moffett datasets, respectively, where the computational time T is also utilized to serve as the metric of computational efficiency. The boldface numbers in these tables indicate the best performance (i.e., the largest RSNR/ERGAS, or the smallest SAM/RMSE) of all fusion algorithms under test. To investigate the performance of image smoothing, the noise levels of HSIs and MSIs data are divided into three cases shown in the ensuing tables.

TABLE 3. Complexity comparison of ADMM iterations for Pavia and Moffett data sets.

Variable	Equation	Complexity	Pavia University		Moffett	
			OM	PRT	OM	PRT
s^{j+1}	(17)	$\mathcal{O}\left(L_h(NL)^2\xi\right)$.	8.577×10^{16}	37.532	8.577×10^{16}	62.487
	(18)	$\mathcal{O}(((Nr^2)^2 + NL)\xi)$	1.259×10^{8}	0.0259	1.259×10^{8}	0.0244
u^{j+1}	(19)	$\mathcal{O}\left((NL)^3)\right)$	8.577×10^{16}	0.357	8.577×10^{16}	0.344
	(20)	$\mathcal{O}\left(N^2L\sqrt{L}\right)$	9.261×10^{8}	0.0839	9.261×10^{8}	0.0807
z^{j+1}	(23)	$\mathcal{O}\left((NL)^3)\right)$	8.577×10^{16}	0.451	8.577×10^{16}	0.452
	(24)	$\mathcal{O}\left(NL\sqrt{L}\right)$	9.261×10^{7}	0.0212	9.261×10^{7}	0.0273
a^{j+1}	(30)	$\mathcal{O}(N^4M^3 + (NM)^2\xi')$	2.037×10^{11}	37.005	1.142×10^{12}	50.635
	(31)	$\mathcal{O}\left(N^4M^3 + N^2L\right)$	1.093×10^{10}	0.0197	6.129×10^{10}	0.0198

For Pavia University dataset in Table 1, one can see that the proposed TVSR-CNMF performs the smallest spectral distortion (the smallest SAM), while the published CO-CNMF algorithm achieves good performance on the other three performance metrics in Case I (i.e. SNR(Ym)=30 dB, SNR(Yh)=25 dB). With the noise power higher, the proposed algorithm outperforms all the baseline algorithms over all four performance metrics in case II (i.e. SNR(Ym)=25 dB, SNR(Yh)=20 dB). Moreover, the proposed TVSR-CNMF has the best performance except for the spectral distortion in the case III (i.e. SNR(Ym)=20 dB, SNR(Yh)=20 dB), i.e. the Bayesian method gets the smallest SAM. In comparison, we can observe that Table 2 for Moffett dataset has the similar trends to Table 1. From the performance tables, we can draw a conclusion that the TVSR-CNMF algorithm can effectively enhance the fusion performance especially under high noise conditions, whereas it may excessively smoothen small scale features of the image, leading to performance degradation in a low-noise or noise free environment.

To evaluate the fusion performance w.r.t. different spectral bands, we also take Case II as an example to display the RSNR and RMSE curves of all fusion algorithms w.r.t spectral bands for two datasets in Fig. 5. One can observe that the proposed TVSR-CNMF algorithm does improve the performance to some extent in comparison with all baseline algorithms over the spectral bands, especially 500-600 nm and 800-850 nm for Pavia University dataset, and 1000-1400 nm for Moffett dataset. These experimental results testify the validity of the total variation and signature-based regularizations on original CNMF criterion (5) and the effectiveness of the carefully designed TVSR-CNMF algorithm.

As for computational efficiency, we can see that the original CNMF has the best performance while the Bayesian method costs the longest running time in Table 1 and 2. In addition, Table 3 shows the complexity comparisons between the naive closed-form solutions and their respective simplified expressions for s^{j+1},u^{j+1}, z^{j+1} and a^{j+1}, which utilize order of magnitude (OM) and per-iteration running time (PRT) as the complexity metrics. In the experiments, (17), (19) and (23) generally run out of memory, due to the large size of this problem. So, some basic processes are employed to yield the results successfully, such as matrix division and Cholesky factorization, etc. However, even so the TVSR-CNMF still takes more than 2000 seconds without

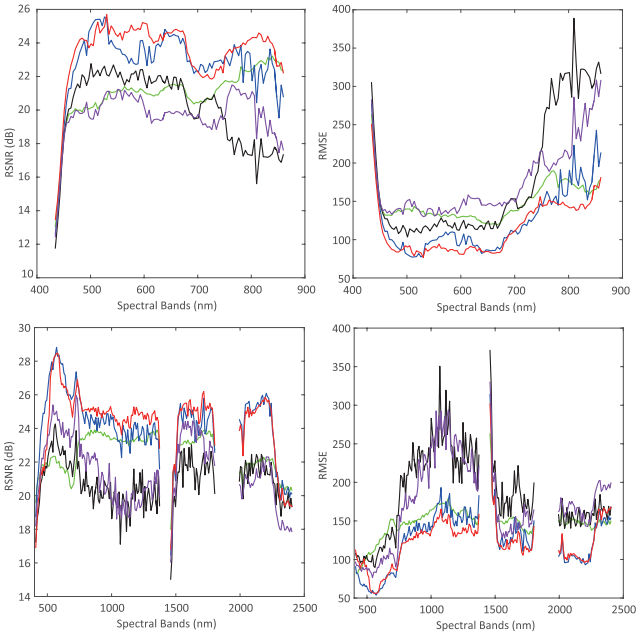

FIGURE 5. The curves of performance (left column) RSNR and (right column) RMSE over spectral bands, including CO-CNMF (blue line), CNMF (black line), Bayesian (green line), HySure (purple line) and the proposed TVSR-CNMF (red line) for (top row) Cuprite and (bottom row) Pavia datasets, respectively.

further complexity reduction. In Table 3, (19) and (23) take much lower PRT than (17) for sparsity, though they have the same OM. After complexity reduction detailed in the appendix, the optimized solutions (18), (20), (24) and (31) greatly reduce the running time, in comparison with their respective naive expressions.

V. CONCLUSION

Incorporating the total variation and signature-based regularizers into the CNMF criterion, we have presented an ADMM-based TVSR-CNMF algorithm for the fusion of hyperspectral and multispectral data, in which only linear closed-form expressions are needed to compute for each ADMM iteration. Our experiments not only demonstrate the superior efficacy of the proposed TVSR-CNMF algorithm over some baseline methods, but also show that the total variation and signature-based regularizations are essential to

enhance the spatial resolution and the quality of the fused data, especially under the environment of high noise level.

APPENDIX

In this part, we detail the procedures of the optimized solutions in Section III, in which some properties and vector-matrix operator of Kronecker product are adopted.

A. PROOF OF VERTICAL-DIFFERENCE-BASED SOLUTION

Due to the large size of vertical difference matrix $\mathbf{D}_v$, (19) is computationally expensive, requiring $\mathcal{O}((NL)^3)$. To make use of the structure of (8), we get

$$\begin{aligned}\mathbf{D}_v^T\mathbf{D}_v &= (\mathbf{I}_{L_x}\otimes\bar{\mathbf{R}}_{L_y}\otimes\mathbf{I}_N)^T(\mathbf{I}_{L_x}\otimes\bar{\mathbf{R}}_{L_y}\otimes\mathbf{I}_N)\\ &= \mathbf{I}_{L_x}\otimes[(\bar{\mathbf{R}}_{L_y}^T\otimes\mathbf{I}_N^T)(\bar{\mathbf{R}}_{L_y}\otimes\mathbf{I}_N)]\\ &= \mathbf{I}_{L_x}\otimes(\bar{\mathbf{R}}_{L_y}^T\bar{\mathbf{R}}_{L_y})\otimes\mathbf{I}_N \qquad (34)\end{aligned}$$

Then, (19) can be expressed as

$$\begin{aligned}\mathbf{u}^{j+1} &= (\mathbf{D}_v^T\mathbf{D}_v+\eta\mathbf{I}_{NL})^{-1}[\mathbf{D}_v^T(\mathbf{v}_1^j+\widetilde{\boldsymbol{h}_2}^j)+\mathbf{s}^{j+1}+\widetilde{\boldsymbol{h}_1}^j]\\ &= (\mathbf{D}_v^T\mathbf{D}_v+\eta\mathbf{I}_{NL})^{-1}\mathbf{D}_v^T(\mathbf{v}_1^j+\widetilde{\boldsymbol{h}_2}^j)\\ &\quad+(\mathbf{D}_v^T\mathbf{D}_v+\eta\mathbf{I}_{NL})^{-1}(\mathbf{s}^{j+1}+\widetilde{\boldsymbol{h}_1}^j)\\ &= [\mathbf{I}_{L_x}\otimes(\bar{\mathbf{R}}_{L_y}^T\bar{\mathbf{R}}_{L_y}+\eta\,\mathbf{I}_{L_y})\otimes\mathbf{I}_N]^{-1}(\mathbf{I}_{L_x}\otimes\bar{\mathbf{R}}_{L_y}^T\otimes\mathbf{I}_N)\tilde{\mathbf{u}}_1\\ &\quad+[\mathbf{I}_{L_x}\otimes(\bar{\mathbf{R}}_{L_y}^T\bar{\mathbf{R}}_{L_y}+\eta\,\mathbf{I}_{L_y})\otimes\mathbf{I}_N]^{-1}\tilde{\mathbf{u}}_2\\ &= \{\mathbf{I}_{L_x}\otimes[(\bar{\mathbf{R}}_{L_y}^T\bar{\mathbf{R}}_{L_y}+\eta\,\mathbf{I}_{L_y})^{-1}\bar{\mathbf{R}}_{L_y}^T]\otimes\mathbf{I}_N\}\tilde{\mathbf{u}}_1\\ &\quad+[\mathbf{I}_{L_x}\otimes(\bar{\mathbf{R}}_{L_y}^T\bar{\mathbf{R}}_{L_y}+\eta\,\mathbf{I}_{L_y})^{-1}\otimes\mathbf{I}_N]\tilde{\mathbf{u}}_2\\ &= \mathbf{vec}(\{[(\bar{\mathbf{R}}_{L_y}^T\bar{\mathbf{R}}_{L_y}+\eta\,\mathbf{I}_{L_y})^{-1}\bar{\mathbf{R}}_{L_y}^T]\otimes\mathbf{I}_N\}\tilde{\mathbf{U}}_1)\\ &\quad+\mathbf{vec}([(\bar{\mathbf{R}}_{L_y}^T\bar{\mathbf{R}}_{L_y}+\eta\,\mathbf{I}_{L_y})^{-1}\otimes\mathbf{I}_N]\tilde{\mathbf{U}}_2), \qquad (35)\end{aligned}$$

whose computational complexity is dominated by two vector-matrix operators. In the first vector-matrix operator, the complexity of $\bar{\mathbf{R}}_{L_y}^T\bar{\mathbf{R}}_{L_y}$ needs $\mathcal{O}(L(\sqrt{L}-1))$, and the inverse of a $\sqrt{L}\times\sqrt{L}$ matrix takes $\mathcal{O}(L\sqrt{L})$. Then, the matrix-matrix product requires $\mathcal{O}(L(\sqrt{L}-1))$. The kronecker product costs $\mathcal{O}(N^2\sqrt{L}(\sqrt{L}-1))$. Finally, the reshaped matrix-matrix product involves $\mathcal{O}(N^2L(\sqrt{L}-1))$. In all, the first operator requires $\mathcal{O}(N^2L\sqrt{L}-1))$. Similarly, the complexity of the second vector-matrix operator is easily to be computed by $\mathcal{O}(N^2L\sqrt{L})$. To sum up, the whole computational complexity of (35) can be given by $\mathcal{O}(N^2L\sqrt{L})$. ■

B. PROOF OF HORIZONTAL-DIFFERENCE-BASED SOLUTION

By the structure of the horizontal difference matrix, (23) can also be simplified to yield the lower complexity. First of all, by means of (9), we have

$$\begin{aligned}\mathbf{D}_h^T\mathbf{D}_h &= (\bar{\mathbf{R}}_{L_x}\otimes\mathbf{I}_{L_y}\otimes\mathbf{I}_N)^T(\bar{\mathbf{R}}_{L_x}\otimes\mathbf{I}_{L_y}\otimes\mathbf{I}_N)\\ &= (\bar{\mathbf{R}}_{L_x}\otimes\mathbf{I}_{NL_y})^T(\bar{\mathbf{R}}_{L_x}\otimes\mathbf{I}_{NL_y})\\ &= (\bar{\mathbf{R}}_{L_x}^T\bar{\mathbf{R}}_{L_x})\otimes\mathbf{I}_{NL_y}. \qquad (36)\end{aligned}$$

Thus, (23) is reformulated by the following procedure as

$$\begin{aligned}\mathbf{z}^{j+1} &= (\mathbf{D}_h^T\mathbf{D}_h+\eta\mathbf{I}_{NL})^{-1}[\mathbf{D}_h^T(\mathbf{v}_2^j+\widetilde{\boldsymbol{h}_4}^j)+\mathbf{s}^{j+1}+\widetilde{\boldsymbol{h}_3}^j]\\ &= (\mathbf{D}_h^T\mathbf{D}_h+\eta\mathbf{I}_{NL})^{-1}\mathbf{D}_h^T(\mathbf{v}_2^j+\widetilde{\boldsymbol{h}_4}^j)\\ &\quad+(\mathbf{D}_h^T\mathbf{D}_h+\eta\mathbf{I}_{NL})^{-1}(\mathbf{s}^{j+1}+\widetilde{\boldsymbol{h}_3}^j)\\ &= [(\bar{\mathbf{R}}_{L_x}^T\bar{\mathbf{R}}_{L_x}+\eta\,\mathbf{I}_{L_x})^{-1}\otimes\mathbf{I}_{NL_y}](\bar{\mathbf{R}}_{L_x}^T\otimes\mathbf{I}_{NL_y})\tilde{\mathbf{z}}_1\\ &\quad+[(\bar{\mathbf{R}}_{L_x}^T\bar{\mathbf{R}}_{L_x}+\eta\,\mathbf{I}_{L_x})^{-1}\otimes\mathbf{I}_{NL_y}]\tilde{\mathbf{z}}_2\\ &= \{[(\bar{\mathbf{R}}_{L_x}^T\bar{\mathbf{R}}_{L_x}+\eta\,\mathbf{I}_{L_x})^{-1}\bar{\mathbf{R}}_{L_x}^T]\otimes\mathbf{I}_{NL_y}\}\tilde{\mathbf{z}}_1\\ &\quad+\{(\bar{\mathbf{R}}_{L_x}^T\bar{\mathbf{R}}_{L_x}+\eta\,\mathbf{I}_{L_x})^{-1}\otimes\mathbf{I}_{NL_y}\}\tilde{\mathbf{z}}_2\\ &= \mathbf{vec}(\widetilde{\mathbf{Z}}_1[(\bar{\mathbf{R}}_{L_x}^T\bar{\mathbf{R}}_{L_x}+\eta\,\mathbf{I}_{L_x})^{-1}\bar{\mathbf{R}}_{L_x}^T]^T)\\ &\quad+\mathbf{vec}(\widetilde{\mathbf{Z}}_2[(\bar{\mathbf{R}}_{L_x}^T\bar{\mathbf{R}}_{L_x}+\eta\,\mathbf{I}_{L_x})^{-1}]^T)\\ &= \mathbf{vec}(\widetilde{\mathbf{Z}}_1\bar{\mathbf{R}}_{L_x}[(\bar{\mathbf{R}}_{L_x}^T\bar{\mathbf{R}}_{L_x}+\eta\,\mathbf{I}_{L_x})^{-1}]^T)\\ &\quad+\mathbf{vec}(\widetilde{\mathbf{Z}}_2[(\bar{\mathbf{R}}_{L_x}^T\bar{\mathbf{R}}_{L_x}+\eta\,\mathbf{I}_{L_x})^{-1}]^T) \qquad (37)\end{aligned}$$

Using the same method as (35),the computational complexity is computed by $\mathcal{O}(NL\sqrt{L})$, due to the facts of $\bar{\mathbf{R}}_{L_y}=\bar{\mathbf{R}}_{L_x}$ and $L_y=L_x=\sqrt{L}$. In comparison, we can observe that the complexity of (37) is less than that of (35) for the difference between (36) and (34). ■

ACKNOWLEDGMENT

The authors would like to thank Prof. Chong-Yung Chi and Dr. Chia-Hsiang Lin from National Tsing Hua University in Taiwan for their guidance and help.

REFERENCES

[1] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. M. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.
[2] J. M. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.
[3] W.-K. Ma *et al.*, "A signal processing perspective on hyperspectral unmixing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 67–81, Jan. 2014.
[4] M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid, and A. Abbas, "Modern trends in hyperspectral image analysis: A review," *IEEE Access*, vol. 6, pp. 14118–14129, 2018.
[5] M. D. Mura, S. Prasad, F. Pacifici, P. Gamba, J. Chanussot, and J. A. Benediktsson, "Challenges and opportunities of multimodality and data fusion in remote sensing," *Proc. IEEE*, vol. 103, no. 9, pp. 1585–1601, Sep. 2015.
[6] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 2, pp. 29–56, Jun. 2017.
[7] L. Loncan *et al.*, "Hyperspectral pansharpening: A review," *IEEE Trans. Geosci. Remote Sens.*, vol. 3, no. 3, pp. 27–46, Sep. 2015.
[8] S. Chaudhuri and K. Kotwal, *Hyperspectral Image Fusion*. New York, NY, USA: Springer, 2013.
[9] Z. Wang, D. Ziou, C. Armenakis, D. Li, and Q. Li, "A comparative analysis of image fusion methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1391–1402, Jun. 2005.
[10] C. Thomas, T. Ranchin, L. Wald, and J. Chanussot, "Synthesis of multispectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1301–1312, May 2008.
[11] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, *25 Years of Pansharpening: A Critical Review and New Developments*. Boca Raton, FL, USA: CRC Press, 2012.

[12] T.-M. Tu, S.-C. Su, H.-C. Shyu, and P. S. Huang, "A new look at IHS-like image fusion methods," *Inf. Fusion*, vol. 2, no. 3, pp. 177–186, Sep. 2001.
[13] P. S. Chavez, Jr., and A. Y. Kwarteng, "Extracting spectral contrast in landsat thematic mapper image data using selective principal component analysis," *Photogramm. Eng. Remote Sens.*, vol. 55, no. 3, pp. 339–348, 1989.
[14] B. Aiazzi, S. Baronti, and M. Selva, "Improving component substitution pansharpening through multivariate regression of MS+Pan data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3230–3239, Oct. 2007.
[15] G. Vivone, R. Restaino, M. Dalla Mura, G. Licciardi, and J. Chanussot, "Contrast and error-based fusion schemes for multispectral image pansharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 5, pp. 930–934, May 2014.
[16] L. Alparone, B. Aiazzi, S. Baronti, and A. Garzelli, "Sharpening of very high resolution images with spectral distortion minimization," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 1, Jul. 2003, pp. 458–460.
[17] J. G. Liu, "Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details," *Int. J. Remote Sens.*, vol. 21, no. 18, pp. 3461–3472, Nov. 2000.
[18] G. Vivone, R. Restaino, G. Licciardi, M. Dalla Mura, and J. Chanussot, "MultiResolution analysis and component substitution techniques for hyperspectral pansharpening," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2014, pp. 2649–2652.
[19] R. C. Hardie, M. T. Eismann, and G. L. Wilson, "MAP estimation for hyperspectral image resolution enhancement using an auxiliary sensor," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1174–1184, Sep. 2004.
[20] Q. Wei, N. Dobigeon, and J. Y. Tourneret, "Bayesian fusion of multi-band images," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1117–1127, Sep. 2015.
[21] Q. Wei, N. Dobigeon, and J. Tourneret, "Fast fusion of multi-band images based on solving a Sylvester equation," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4109–4121, Nov. 2015.
[22] M. Simões, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3373–3388, Jun. 2015.
[23] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, Feb. 2012.
[24] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.
[25] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J. Y. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3658–3668, Jul. 2015.
[26] Z. H. Nezhad, A. Karami, R. Heylen, and P. Scheunders, "Fusion of hyperspectral and multispectral images using spectral unmixing and sparse coding," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 2377–2389, Jun. 2016.
[27] E. Wycoff, T.-H. Chan, K. Jia, W.-K. Ma, and Y. Ma, "A non-negative sparse promoting algorithm for high resolution hyperspectral imaging," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2013, pp. 1409–1413.
[28] M. D. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 3, pp. 542–552, May 1994.
[29] C.-H. Lin, W.-K. Ma, W.-C. Li, C.-Y. Chi, and A. Ambikapathi, "Identifiability of the simplex volume minimization criterion for blind hyperspectral unmixing: The no-pure-pixel case," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5530–5546, Oct. 2015.
[30] Y. Zhang, Y. Wang, Y. Liu, C. Zhang, M. He, and S. Mei, "Hyperspectral and multispectral image fusion using CNMF with minimum endmember simplex volume and abundance sparsity constraints," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2015, pp. 1929–1932.
[31] C.-H. Lin, F. Ma, C.-Y. Chi, and C.-H. Hsieh, "A convex optimization-based coupled nonnegative matrix factorization algorithm for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1652–1667, Mar. 2018.
[32] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images," *Photogramm. Eng. Remote Sens.*, vol. 63, no. 6, pp. 691–699, 1997.
[33] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[34] C.-H. Lin, C.-Y. Chi, Y.-H. Wang, and T.-H. Chan, "A fast hyperplane-based minimum-volume enclosing simplex algorithm for blind hyperspectral unmixing," *IEEE Trans. Signal Process.*, vol. 64, no. 8, pp. 1946–1961, Apr. 2016.
[35] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1327–1344, Oct. 2004.
[36] M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Dunne, and J. F. Huntington, "ICE: A statistical approach to identifying endmembers in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2085–2095, Oct. 2004.
[37] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4484–4502, Nov. 2012.
[38] C.-Y. Chi, W.-C. Li, and C.-H. Lin, *Convex Optimization for Signal Processing and Communications: From Fundamentals to Applications*. Boca Raton, FL, USA: CRC Press, 2017.
[39] J. Sigurdsson, M. O. Ulfarsson, and J. R. Sveinsson, "Blind hyperspectral unmixing using total variation and ℓ_q sparse regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6371–6384, Nov. 2016.
[40] S. Arora *et al.* (2012). "A practical algorithm for topic modeling with provable guarantees." [Online]. Available: https://arxiv.org/abs/1212.4777
[41] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.
[42] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Upper Saddle River, NJ, USA: Prentice-Hall, 1989.
[43] *ROSIS Free Pavia University Data*. Accessed: Nov. 1, 2017. [Online]. Available: http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes
[44] *IKONOS IKONOS Satellite Imagery*. Accessed: Nov. 1, 2017. [Online]. Available: https://apollomapping.com/imagery/high-resolution-imagery/ikonos
[45] *AVIRIS Free Standard Data Products*. Accessed: Nov. 1, 2017. [Online]. Available: http://aviris.jpl.nasa.gov/html/aviris.freedata.html
[46] M. A. Bendoumi, M. He, and S. Mei, "Hyperspectral image resolution enhancement using high-resolution multispectral image based on spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6574–6583, Oct. 2014.
[47] Q. Wei, J. Bioucas-Dias, N. Dobigeon, J.-Y. Tourneret, M. Chen, and S. Godsill, "Multiband image fusion based on spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7236–7249, Dec. 2016.
[48] T. Stathaki, *Image Fusion: Algorithms and Applications*. Oxford, U.K.: Academic, 2008.
[49] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2006.
[50] C.-I Chang, "A review of virtual dimensionality for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1285–1305, Apr. 2018.
[51] C.-I. Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, Mar. 2004.

FEIXIA YANG received the B.S. and M.S. degrees in control theory and engineering from Liaoning Technical University, Fuxin, China, in 2003 and 2006, respectively. She is currently pursuing the Ph.D. degree in electronic science and technology with the Beijing University of Posts and Telecommunications, Beijing, China. She has been a Faculty Member with the School of Electrical and Control Engineering, Liaoning Technical University, Huludao, China, since 2006. Her research interests are in digital image processing, remote sensing imaging and pattern recognition.

FEI MA received the B.S. degree in communication engineering and the M.S. degree in computer science from Liaoning Technical University, Fuxin, China, in 2001 and 2004, respectively, and the Ph.D. degree in circuit and system from the Beijing University of Posts and Telecommunications, Beijing, China, in 2010. He held a post-doctoral position at the University of Electronic Science and Technology of China, Chengdu, China, in 2014. Then, he conducted research as a Visiting Scholar with National Tsing Hua University, Hsinchu, Taiwan, from 2015 to 2016. He is currently a Faculty Member with the School of Electronic and Information Engineering, Liaoning Technical University, Huludao, China. His research interests include hyperspectral image processing, radar signal processing, pattern recognition, and convex optimization.

Ziliang Ping received the B.S. degree in physics from the Nankai University of China in 1968 and the M.S. degree in photo-electrics from the Beijing University of Posts and Telecommunications of China in 1983. He is currently a Professor with the Beijing University of Posts and Telecommunications. He has published about 100 papers in many journals and international conferences, such as *Journal of the Optical Society of America* and*Pattern Recognition*. So far, he has completed four national natural sciences funds and six provincial natural science funds. He has been enjoying the special dedication allowance issued by the Chinese Government since 1993. His research interests involve digital image processing, optical information processing, and computer vision.

GUIXIAN XU received the Ph.D. degree in communication and information systems from the Beijing University of Posts and Telecommunications, Beijing, China, in 2018. Since 2011, he has been with Datang Mobile, Beijing, as a TD-LTE Hardware Test Engineer. He was an Intern with CBC/XEV of Ericsson China and the State Key Laboratory of Wireless Mobile Communications, China Academy of Telecommunications Technology, Beijing. He was a Visiting Ph.D. Student with National Tsing Hua University, Hsinchu, Taiwan, from 2015 to 2016. He currently holds a post-doctoral position at the Connectivity Section, Department of Electronic Systems, Aalborg University, Denmark. His research interests are in massive MIMO, 5G, and convex optimization in wireless communications, and signal processing.